



Tatton-Brown, O. M. W. (2019). Rigour and Intuition. *Erkenntnis*.
<https://doi.org/10.1007/s10670-019-00180-9>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1007/s10670-019-00180-9](https://doi.org/10.1007/s10670-019-00180-9)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Springer Nature at <https://link.springer.com/article/10.1007/s10670-019-00180-9>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



Rigour and Intuition

Oliver Tatton-Brown¹

Received: 10 June 2018 / Accepted: 3 October 2019
© The Author(s) 2019

Abstract

This paper sketches an account of the standard of acceptable proof in mathematics—rigour—arguing that the key requirement of rigour in mathematics is that non-trivial inferences be provable in greater detail. This account is contrasted with a recent perspective put forward by De Toffoli and Giardino, who base their claims on a case study of an argument from knot theory. I argue that De Toffoli and Giardino’s conclusions are not supported by the case study they present, which instead is a very good illustration of the kind of view of proof defended here.

Recently there has been much philosophical discussion of what the canons of correctness in mathematics are. For a while philosophers of mathematics implicitly assumed what has been dubbed the “standard view of proof”, in which the correctness of mathematical proofs is connected with their formalizability in some way. A wide range of authors have recently criticized this view, arguing that the connection with formalizability is spurious and is imposed on mathematics from the outside by logicians and philosophers. For example, Rav (1999), Rav (2007), Celluci (2009), Leitgeb (2009), Pelc (2009), Goethe and Friend (2010), Antonutti Marfori (2010), Larvor (2012, 2019), De Toffoli and Giardino (2016), and Weir (2016). Regrettably there does not seem to be a paradigmatic statement of the standard view of proof in print, and these authors generally criticize what they perceive to be the consensus without citing a specific target position. Azzouni (2013) and Burgess (2015) have put forward versions of the standard view of proof, but neither really engages with the critics, or has attracted much attention from them.

My aim here is limited: to briefly sketch a version of something like the standard view of proof, and contrast it with the perspective that seems to emerge from De Toffoli and Giardino (2016). They aim to use a case study to attack a version of the standard view of proof, and motivate an alternative position. I discuss the

Supported by AHRC grant AH/L503939/1.

✉ Oliver Tatton-Brown
o.tatton-brown@bristol.ac.uk

¹ Department of Philosophy, University of Bristol, Cotham House, Bristol BS6 6JL, UK

conclusions they draw from their case study, agreeing with some, but arguing that the bolder conclusions are based on two misunderstandings of the case study they present. I conclude that the case study is no problem for the account of proof I defend—in fact it is a good illustration of it.

1 Rigour

I will start by discussing the standard for acceptable proof in much of modern mathematics: rigour. Regrettably this discussion will have to be brief. I give a more detailed account of rigour in another paper which I will advert to in places,¹ but for now hopefully the essentials will suffice. In places I will draw on points made by Burgess (2015).

One can start characterizing rigour by giving examples of what it is not. The kinds of naive manipulations involving infinitesimals made in the seventeenth and eighteenth centuries were not generally rigorous, lacking clear rules for what kinds of reasoning were and weren't acceptable. Euler's famous solution of the Basel problem, by assuming the sine function is like a finite polynomial and can be factorized in terms of its roots, is another example of non rigorous reasoning (Dunham 1999, pp. 45–48). His argument was simple and ingenious and gave an answer whose rough correctness could be checked by calculation, but his factorization of the sine function in terms of its roots could not (at that point) be proved correct.

There are a couple of lessons which can be immediately drawn from these kinds of simple examples. Firstly, the modern standard of rigour is different to prior standards of acceptable proof. Secondly, rigour does not just mean “reliable reasoning” or “explanatory reasoning” or something along these lines. In the right hands, naive calculations involving infinitesimals could be perfectly reliable and explanatory, and Euler's solution to the Basel problem likewise; yet these are not rigorous arguments. Thirdly, mathematics does not have to be rigorous to be valuable. The kinds of arguments described above were worth giving, and the same goes for mathematical arguments made by modern physicists, engineers and so on.

Though non rigorous mathematics can be reliable, explanatory, and valuable, it is still worth giving attention to rigour as a standard. Modern rigorous mathematics is after all mathematics at its deepest and most fruitful. In another paper, mentioned above² I argue that restricting attention to rigorous mathematics allows us to give a relatively straightforward epistemology (modulo the important question of the justification of the basic principles).

So can more be said about what rigour consists in than the negative characterization given above? The basic perspective here is that rigour amounts to the ability to prove statements in greater detail. This is only a very rough characterization, and needs to be clarified in various ways.

¹ “Rigour and Proof”, under review.

² “Rigour and Proof”, under review.

Firstly, this makes it sound like one could keep on increasing the detail of an argument forever. In reality, one can make inferences in a proof which are already as detailed as possible. If an inference has been directly justified in previous arguments, or is an instance of a basic principle, then there is no more detail to add (beyond perhaps a citation, but that is not a part of the proof itself). Burgess has emphasised that from the point of view of producing new mathematics, it does not make much difference whether one appeals to a basic principle or an existing result, as long as the existing result was itself rigorously deduced: in this sense which the basic principles are taken to be is often unimportant (Burgess 2015, pp. 149–158).

It will be worth saying a little bit more about the basic principles, and their relation to individual branches of mathematics. It is very common in mathematics for results from one area to be fruitfully applied in a different area. Burgess emphasises that part of making this rigorous, and ensuring the compatibility of the different branches, consisted in finding a single set of basic principles from which all the different branches could be deduced (Burgess 2015, pp. 56–63). When developing a new branch of mathematics, one does not introduce new basic principles for it: one shows how to define its objects in terms of objects already known, and one derives their basic properties as required in the usual way. For instance associativity is in a sense an axiom of group theory, but we do not need to posit it as a new basic principle—it is just a property that (by definition) any group has.

There are occasionally what might look like exceptions to this, notably the axiom of universes sometimes used when working with categories (in particular in modern algebraic geometry, following Grothendieck). This is not an ad hoc assumption about categories however. One can be justified in appealing to it because (the feeling goes) it could perfectly well have been amongst the basic principles from the start. It can be precisely stated, can be motivated philosophically in a similar way to the other axioms, and is known to be independent of them.

Having noted that special case it can be put to one side. The point so far is that rigorous proofs take place in a context, consisting of background facts and inferences already justified, which can be appealed to when needed.

The context will also contain existing concepts, that can be employed when reasoning or when forming new definitions. As Burgess mentions (Burgess 2015, p. 7), in rigorous mathematics new concepts have to be introduced by a clear definition in terms of existing ones. A definition does not have to be completely formal: for instance one might describe a topological space as a set X equipped with a topology \mathcal{T} , without specifying whether this means a pair (X, \mathcal{T}) or a pair (\mathcal{T}, X) or something else. However it does have to be clear that a definition could be made precise in such a way that the argument would be valid.

However proofs do not generally consist of just chaining together inferences already directly justified. An argument like that would be one written out in maximum detail and really, there are many different levels of detail that rigorous mathematics can take place at (here we are entering territory Burgess does not cover).

To discuss this, we can start at the beginning. Real analysis is commonly used as the subject in which students first learn serious mathematics in a rigorous way, and at the start of a real analysis course students are often presented with arguments at a very great level of detail. This can be seen in a standard analysis text like Abbott

(2016): in sections 1.2, 1.3 and the early parts of 1.4 (as far as Theorem 1.4.3), the proofs are mostly about as detailed as one could make them. Pretty much all of the logical structure of the arguments is right there on the page: for instance the demonstration of existential and universal statements is carried out in a way that closely parallels the corresponding natural deduction rules. It would not be a challenge to formalize these arguments. We can describe the level of detail these arguments take place at as the “week 2 level of detail”—where arguments are strung together out of these very basic, detailed inferences. Of course a student might not actually work through these examples in week 2 (or at all), this is just a convenient name. By studying and writing these kinds of arguments, students will hopefully learn what we can call “proficiency at the week 2 level of detail”, the ability to prove simple facts in this basic level of detail.

Mathematical arguments gradually get faster and higher level as a student advances. By Chapter 7 of Abbott (2016), the proofs are a bit less detailed: some of the manipulations of partitions and limits take place at a slightly higher level, with less attention paid to justifying claims directly in terms of the definitions. We can call the rough level of detail of arguments in this chapter the “term 2 level of detail” (this is again just a convenient name). Arguments at the term 2 level of detail might not proceed as explicitly in terms of the definitions as at the week 2 level of detail, but it is clear that every inference in one of these arguments could be spelled out at the week 2 level of detail if required.

One reason that students first learn proficiency at the week 2 level of detail is that then when confronted with slightly higher level arguments, they can judge for themselves how plausible the inferences are. They will hopefully have gained a reliable sense for how concepts like limits behave, and can then accurately judge some inferences involving them to be correct without having to consult the definitions; and in cases where a higher level inference does not seem immediately obvious, a student can see if they can come up with an argument to justify it, descending when needed to the level of greater detail they have already mastered.

One can keep picking out other levels of detail in the same way: one could give some examples of arguments at a slightly higher level still, and call it the “year 2 level of detail”, then define a “year three level of detail” and so on. These terms will of course be somewhat vague and subjective, in the same way as with almost any other concept we use (such as “red” or “chair”). Again it should be emphasised that the terms “year 2”, “year 3” and so on are just convenient names, and it is not being assumed that all mathematics in year 3 takes place at exactly the same level of detail. Even when teaching the same subject at the same stage there can be a choice about how much detail to include. For instance the textbooks Hirsch (1976) and Lee (2012) both cover differential topology at the graduate level, with considerable overlap—in chapters 1, 3 and 4 of Hirsch and chapters 1, 2, 3, 4, 5, 6 and 10 of Lee—but Hirsch’s proofs are often terser. One could use arguments from Lee as examples of a “graduate level of detail (explicit)”, and arguments from Hirsch as examples of a “graduate level of detail (terse)”.

The talk of “detail” here means explicitness, proximity to definitions, there not being much more that could be added to an argument. This is not the same as judging the complexity of the concepts involved. Smooth manifolds are reasonably

complex as mathematical concepts go, but one can still reason about them in a very detailed way—as seen in Lee (2012, Proposition 2.4).

Students will proceed through these levels of detail as they learn mathematics, gradually being exposed to higher level, less detailed arguments over time. At every stage, they can (hopefully) use their proficiency at a given level of detail to help when it comes to grasping and forming less detailed arguments: their knowledge of the concepts involved will be grounded in an ability to use them, to prove facts involving them, and if their judgement is ever unsure about a high level inference then they can try to prove it at the level of detail they are already comfortable with, sharpening their judgement in the process. Implicit in this is that it is constitutive of an argument being valid at the level of greater compression that each inference can be spelled out at the level of greater detail. Above it was discussed how the arguments in Hirsch (1976) are often terser than those in Lee (2012). This is not a problem; but if there were an inference in Hirsch that could not be carried out at the level of detail of Lee, no matter how hard one tried, this definitely would be a sign that something was wrong with it.

The focus so far has been on detail, and the ability to prove things in greater detail, but this is not the whole story. One can (rightly) see a higher level piece of reasoning to be correct without thinking through a more detailed justification, and this is how arguments will often be read and written. The requirement of rigour is that the ability to prove things in greater detail is always there, if necessary. This is cited in the Princeton Companion to Mathematics as an important mechanism for resolving disputes about the correctness of proofs (Gowers et al. 2008, p. 74). It is also an important as a way for a mathematician to train and refine their high level judgements whenever necessary, by seeing kinds of which high level judgements can be backed up by a proof.

There may be cases where one could use the term “intuition” to describe certain higher level mathematical judgements, particularly where the reasoning is in some sense spatial or temporal. For the judgements to be valid in a rigorous proof though, this will have to be intuition of a rather special kind. One cannot just be giving an untutored judgement of the plausibility of a claim: one has to be judging its provability. A classic example to illustrate this is the Jordan curve theorem, which states roughly that every continuous injective closed curve in the plane has an inside and an outside. This is intuitively about as obvious a statement as one can give, and to someone without experience of pathological functions it is probably hard to imagine what a counterexample could possibly look like. Nonetheless the proof is famously hard (if one works from the definitions, without tools like algebraic topology). Part of learning rigorous mathematics is learning to tell the difference between a statement like the Jordan curve theorem—which is obvious, but hard to prove—and a statement like the intermediate value theorem, which is obvious and whose proof is in fact straightforward. Of course intuitive judgements of plausibility are very important in mathematics: it is crucial that us humans are able to judge a statement like that of the Jordan curve theorem to be very likely, and thus set out to prove it. But when doing rigorous mathematics, there is a great difference between the kinds of judgements that would guide research in this way, and the kinds of judgements that are acceptable in a proof itself.

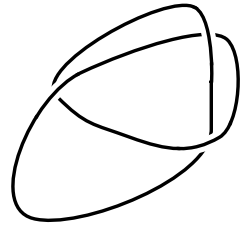
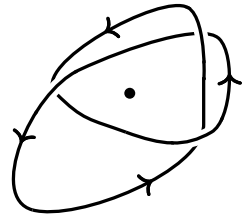
To finish the section I will sketch why on this view valid proofs are in fact formalizable (in a sense). Thus the account here can be seen as a version of the standard view or proof, as discussed in the introduction. We argue this by an induction upwards through levels of detail, starting with the week 2 level of detail and going up and up towards the research frontier. It was noted above that just by looking at the kinds of examples given of arguments at the week 2 level of detail, it is clear that they are formalizable. Then any inference at the term 2 level of detail can be proved at the week 2 level of detail—as noted above—and thus can be proved formally. Thus any argument at the week 2 level of detail is made up of formalizable inferences, so is formalizable. Then we just keep going in this way. Any inference at the year 2 level of detail can be proved at the term 2 level of detail, so is formalizable. Thus any argument at the year 2 level of detail consists of formalizable inferences, so is formalizable (in principle). At every stage for an inference to be valid at a higher level of detail, it is necessary that it be provable at one notch greater detail; thus we can keep going up and up through levels of less and less detail, arguing that proofs at each level are formalizable. One can reach the sketchiest level of detail acceptable at the research frontier in a small finite number of such steps up, and thus we obtain that all valid proofs are (in principle) formalizable.

This is quite a quick argument, and there are ways it could be clarified, and objections that could be considered. My intention here is just to sketch why there is a connection between validity and formalizability on this account of rigour.³ I am not claiming that formalized proofs are more convincing for humans than high level ones, that there is a unique right way to formalize a given high level proof, that a formalization shows the real reason that a proof is correct, or anything else along these lines. These are dubious claims that critics of the standard view of proof rightly criticize (though I am not sure whether claims like these are actually widely believed).

2 An Alternative Perspective

I will now describe the perspective of De Toffoli and Giardino (2016). They give general comments on mathematical proof, basing them around a case study: a result from knot theory known as Alexander's lemma. Here I will briefly describe Alexander's result, and discuss the general view put forward by De Toffoli and Giardino. Then in Sect. 3 I relate De Toffoli and Giardino's version of Alexander's argument. As we will see in Sects. 4 and 5, some key aspects of De Toffoli and Giardino's version—used to support their central conclusions—are not in fact found in Alexander's original argument. These aspects are introduced in De Toffoli and Giardino's retelling, partly because it combines together aspects of both Alexander's original version and Jones's (1998).

³ As mentioned above, more detail on this, and the rest of the above account of rigour, is found in my paper "Rigour and Proof", under review.

Fig. 1 A knot diagram**Fig. 2** A knot diagram winding around an axis

Alexander's result concerns knots, which for now can be thought of roughly as loops of string in space. The exact definition is actually relevant to Alexander's argument, and De Toffoli and Giardino's analysis of it, as will be seen later in Sect. 4. A key tool in knot theory is the ability to project a knot on a suitable plane, obtaining a knot diagram that indicates all its salient features (Fig. 1).

Alexander's lemma states that every knot is equivalent to one with a diagram that only winds one way around an axis. Figure 2 shows this to be true of the diagram in Fig. 1.

We will put the proof of this result to one side for now, returning to it in Sects. 3 and 5 when discussing certain claims De Toffoli and Giardino make about it.

On to De Toffoli and Giardino's account. They are partly motivated by the call from Larvor (2012, p. 716) for philosophers of mathematical practice to develop better answers to the questions of "What is the philosophy of mathematical practice?" and "How does one do it?". Larvor finds the standard answers—involving an aspiration to study 'actual' mathematical activity, and complaints about other approaches to philosophy of mathematics which assume the validity of formal models of proof—unsatisfying, and De Toffoli and Giardino (2016, pp. 26–27) agree.

In answer to the first question, they propose defining philosophy of mathematical practice as the analysis of mathematicians' use of representations (De Toffoli and Giardino 2016, p. 27). I will not pause too long to quibble with this, but there are two remarks worth making. Firstly, if by representation one means a visual representation of an object that one can manipulate (as with representations of knots, Toffoli and Giardino's topic in this paper), then restricting one's focus this narrowly will miss out large portions of mathematics. Indeed it is not uncommon in mathematics for reasoning to proceed (for instance) by symbolic manipulations, using a tutored instinct for where the manipulations can lead rather than the guidance of a visual representation—perhaps best illustrated by algebra, and the study of groups, rings, field extensions, co-algebras, chain complexes and so on (even here they may often be a certain amount of visualization accompanying reasoning, but this is not

generally the grounds for the inferences being made). Secondly, I would argue that when it comes to rigorous mathematics involving visualization, the key focus should not be on visual representations and what appears to be true of them, but on when and how one can reason in high level, intuitive ways using visual representations and still reliably judge that one's inferences could be backed up by a proof if required. It was argued in Sect. 1 that the ability to back inferences up with more detailed proofs if necessary is the key feature of rigorous mathematics, and the Jordan curve theorem was given as an example of a statement that seems obviously true based on one's visual representation of the situation, but which nevertheless cannot be rigorously asserted without proof. Whether Alexander's lemma fits this rigorous paradigm—an intuitive argument whose inferences can be backed up by proofs—will be one subject of this paper.

Another of De Toffoli and Giardino's main aims (again following Larvor 2012, p. 716) is to

challenge the model of formal logic as adequate to account for proof (De Toffoli and Giardino 2016, p. 27)

It is not totally clear what view they (and Larvor) are intending to counter here, however. Formal proofs are a model of proof, and how good or bad a model is will depend on what you are using it for. It is not clear whether anyone has claimed that formal proofs are the right model in all circumstances—this is certainly not claimed by Azzouni (2013) or Burgess (2015), two recent defenders of a link between informal and formal proofs. There are some ways in which formal proofs are obviously unlike the proofs mathematicians write: if all you know is the rules of natural deduction, there is no way you will be able to follow research level mathematics, no matter how smart you are. On the other hand there are also ways in which formal proofs are a good model of proof—for instance if a formal independence result is discovered, showing that some statement cannot be formally proved or disproved from the axioms of ZFC, then there is no point in trying to find an informal proof using normal mathematical reasoning—one would have to try adding some additional basic principle.

However it is definitely true that there is more to say about informal proof than just that it is modelled by formal proof. A simple account of how the standard of proof in much of mathematics—rigour—works was seen in Sect. 1. Some of what De Toffoli and Giardino go on to say is compatible with that, and can be seen as useful additions to it for the particular case of low dimensional topology, or more widely. Some of their account is in contradiction with it though, and I will generally take issue with these parts. Some of their stronger claims in this direction are unsupported by Alexander's proof, as we will see in Sects. 4 and 5.

De Toffoli and Giardino are right to emphasise the collective aspects of mathematical practice (De Toffoli and Giardino 2016, pp. 28–29). They are also right to emphasise the harnessing of existing human cognitive capacities during mathematical reasoning (De Toffoli and Giardino 2016, pp. 29–30). This would be included under what I vaguely termed “high level reasoning” in Sect. 1. However—as mentioned above—I would amend their discussion to emphasise that when it comes to rigorous mathematics, the important question is how the

existing cognitive capacities are linked with judgements of provability: how does one learn reliably that certain natural ways of reasoning can be backed up (given the time and inclination) with proofs? They next discuss representations, in particular systems of notation, and I think they make important points here about the value of efficient, suggestive notation (De Toffoli and Giardino 2016, pp. 30–32).

De Toffoli and Giardino's next topic of "permissible actions" is the main one about which I have reservations. The concept is drawn from Larvor (2012). To motivate their discussion of permissible actions, they appeal to a quote from Jones (1998):

I remember being worried by Russell's paradox as a youngster, and am still worried by it, but I hope to demonstrate ... that it is not at all difficult to live with that worry while having complete confidence in one's mathematics (Jones 1998, p. 203; De Toffoli and Giardino 2016, p. 203)

They infer from this quote that confidence in mathematics is not based on "logic' or foundations", and ask what the actual grounds for conviction are. It is worth saying a bit about this before moving on to discuss permissible actions. A basic point is that it is crucial to distinguish how one can gain *conviction* in mathematics from the question of what the standard of *proof* is in mathematics. It is certainly true that rigorous proof is not the only way to gain conviction in mathematics: indeed this is the point of Jones's quote above, and he supports it with a number examples, such as his discussion of the huge number of applications of the Fourier transform, making the point that even if all our proofs of its properties turned out to be fallacious (or built on inconsistent assumptions) there must still be some sense in which this transform is true or valid (Jones 1998, pp. 203–204). In Sect. 1 we also saw a number of examples of inferences that may be convincing (and some of which were once accepted as valid), but would not be acceptable by the modern standard of rigour in mathematics. Though Jones is correct that conviction can be generated without a rigorous proof, that does not mean that that is the norm in mathematics, or that we should look elsewhere for the actual grounds for conviction; nor is this evidence either for or against any analysis of proof, whether based on logic or otherwise.

Now, onto the topic of permissible actions. De Toffoli and Giardino believe that mathematicians can gauge whether a proof is correct by seeing whether it consists entirely of these permissible actions, which are ways of reasoning that are accepted by the community of practitioners. As they put it,

To become a practitioner means to learn to operate correctly on the representations, that is, to perform the appropriate actions. (De Toffoli and Giardino 2016, pp. 32–33)

They describe the proof as being addressed to this particular community of practitioners, a community which

defines the 'permissible actions' on the representations. (De Toffoli and Giardino 2016, p. 44)

They believe that when Alexander refers to “legitimate operations” he means these kinds of “permissible actions”. They describe these as

part of [the community’s] mental model, [which] can be considered as reliable to gain new knowledge about the object of research. (De Toffoli and Giardino 2016, p. 45)

They also introduce the term “local criteria of validity” in this connection, arguing that different areas of mathematics will have “different criteria of validity” (De Toffoli and Giardino 2016, p. 49).

It is true that there are standards for what is acceptable in mathematical proof, and I would agree that there is no logic based criterion for this. An attempt to roughly describe how the standard of rigour in mathematics works was given in Sect. 1, and I did not put forward a criterion: one with no prior experience of evaluating mathematics could not read that account and hope to be able to judge the correctness of proofs. The distinctive feature of De Toffoli and Giardino’s analysis is in seeing mathematicians as split into distinct communities, each with their own idiosyncratic ways of reasoning and their own standards of correctness—standards which each individual community defines, without any further justification being supplied or called for. De Toffoli and Giardino write as though each community’s ways of reasoning are automatically accurate about the community’s chosen subject matter, because they form part of the mental model the community shares.

One obvious question this analysis ignores is where these communities come from. Practitioners of the various branches of mathematics have not been passing their wisdom down from one generation to the next since time immemorial. Most branches of modern mathematics have only existed in their present form since around 1900 or later, with the modern notion of mathematical rigour only stemming from around that time. It is not clear how the creation of new branches of mathematics and new mathematical communities would fit into De Toffoli and Giardino’s account. They seem to be denying any general standard for what is acceptable in proof, which suggests that each community is free to set its standards as it likes on formation (though as De Toffoli and Giardino tell it, these standards appear to be fixed once they have been accepted by the community). Can any group of people studying mathematical subject matter call themselves a community of mathematicians, no matter how they do it? What if they extend the notion of proof to include numerical evidence, or conclusions reached in dreams?

This is obviously silly, and the reality is that the creation of new branches of mathematics is a routine part of the ordinary functioning of the subject. Indeed new branches of mathematics—studied by a particular “community”—are invented with some regularity. As discussed in Sect. 1, one cannot just make up whatever kind of mathematics one likes, positing the existence of new kinds of objects, and hypothesising ways in which they behave: in rigorous mathematics, the birth of a new branch requires a demonstration of how its objects can be defined in terms of existing concepts, and how its basic principles can be demonstrated as consequences of these definitions. Witness the rigorous development of probability by Kolmogorov in terms of sigma algebras, the rephrasing by Grothendieck of algebraic geometry

in terms of schemes, the development by Voevodsky of motivic cohomology, and many other examples.

It is true that in each branch there will be distinctive ways of reasoning, or “permissible actions”. However De Toffoli and Giardino appear to suggest that these are reliable because the community accepts them, so they form part of the shared mental model of the practitioners, and thus part of the subject matter of the branch. In reality, in rigorous mathematics the opposite is the case. The permissible actions are not reliable because the community accepts them: the community accepts them because they are reliable—because they can be seen and checked to be accurate ways of reasoning about the subject matter, according to the definitions given.

As well as being too permissive in its implications for what standards a community of mathematicians can set, the analysis in terms of permissible actions also does not properly reflect the pervasiveness and importance of novelty in mathematical arguments. De Toffoli and Giardino do accept the possibility that the practice of mathematics may evolve, for instance with material representations (symbols, notation, diagrams and so on) stemming from certain mental models, but then leading to insights which feed back in and modify the mental models themselves (De Toffoli and Giardino 2016, p. 30). But this is only a potential source of gradual change in the standard of proof that a community accepts, and it seems that at each point in time on this view there is still a fixed list “permissible actions” which states what kinds of inferences can be made, a list taught to each new practitioner as a student. If a new kind of argument is made, not comprised of inferences on the list of permissible actions, then whether this argument is valid or not will (apparently) come down to whether the community can be persuaded to change their standards of proof to accept it.

In reality however mathematical reasoning is not nearly so constricted. Consider the introduction of probabilistic methods into combinatorics by Erdős, the application of linear algebra to group theory by Frobenius and others, and the development of homology by numerous mathematicians (including Alexander himself). If a brilliant mathematician develops a new way of reasoning about some object, then if that way of reasoning is correct, and can be seen to be correct, and justified in greater detail and precision if necessary, then it is a valid way of reasoning—even if the community had never even considered it before. Many breakthroughs in mathematics consist of exactly this. Even in more ‘everyday’ mathematics, papers will often contain new ways of arguing and new ideas, but on a smaller scale. The novelty we see in mathematics is possible precisely because there is a general standard for acceptable proof, one not constituted by the methods each community of mathematicians currently happens to use.

Instead of supporting De Toffoli and Giardino’s analysis, Alexander’s result actually turns out to be a good illustration of where it goes wrong. De Toffoli and Giardino hope to argue for mathematics as split up into separate communities, with their own standards of proof and ways of reasoning, but it is not even clear whether there was an established community of knot theorists at the time Alexander was writing (1923): this was before some of the major inaugural results of the field, such as Reidemeister’s theorem—the seminal theorem which states that any equivalent knots have diagrams which can be related by a finite sequence of the

three Reidemeister moves (Alexander and Briggs 1926; Reidemeister 1927). At any rate Alexander's paper is not addressed to such a community, using ways of reasoning that only an initiate would understand. On the contrary, the entire argument is elementary, and straightforward to anyone with a basic knowledge of mathematics. He does not assume a background knowledge of knot theory, gesturing at a simple definition of knot as composed of a finite number of straight pieces (though he does not state this completely precisely)—here we see the concepts of the new field being defined in terms in existing concepts, as discussed above and in Sect. 1. Based on this definition, we can follow the argument, and see his inferences about knots to be accurate—not because we have been taught special kinds of reasoning used by knot theorists, but because we already have a grasp of how straight line segments in \mathbb{R}^3 behave. We can follow the argument involving the new concepts because of our grasp of the existing concepts, checking any inferences in greater detail as necessary. The argument Alexander gives is rigorous by the general standard used throughout mathematics, discussed in Sect. 1—it does not rely on some special standard of proof used by knot theorists.

It is true that one key concept in the argument—the “legitimate operations”—goes undefined, but it is clear from the context what this is intended to mean, as I discuss in Sect. 4. This is one critical point where De Toffoli and Giardino misinterpret Alexander, apparently taking him to be working with an intuitive notion of continuous (or smooth) transformation, without precise definition. This misinterpretation leads them to misinterpret the inference Alexander makes with the notion, which leads them in turn to overstate Alexander's reliance on intuition and visualization.

A second respect in which they misinterpret Alexander, also leading them to overstate his reliance on intuition and visualization, is in the structure of his argument: how his argument ensures that the process of knot modifications described terminates. Again they claim he is relying purely on intuition to justify this, and again their claim is erroneous (as a claim about Alexander's argument), as I discuss in Sect. 5.

First, I will briefly describe De Toffoli and Giardino's account of Alexander's argument, before looking at these two aspects in detail.

3 De Toffoli and Giardino's Account of Alexander's Argument

There are currently three versions of Alexander's argument in play: Alexander's original proof (Alexander 1923), a description by Field's medallist Vaughan Jones in a philosophical piece (Jones 1998), and the version of De Toffoli and Giardino in their own philosophical piece (De Toffoli and Giardino 2016). Alexander's and Jones' versions are importantly different, but De Toffoli and Giardino's version combines together aspects of both, and this is where the problems stem from.

Here I will limit myself to describing the key features of De Toffoli and Giardino's version. If one wants to see the actual proof it is best to look at the original, which is brief, simple and clearly written (Alexander 1923).

It will be relevant to the following that there are different notions of knot one can work with. A polygonal knot is made up of a finite number of straight line segments, intersecting only at their endpoints. A smooth knot is a smooth non self intersecting map $S^1 \rightarrow \mathbb{R}^3$.

We also need the notion of equivalence of knots, which again can be defined in different ways. The general notion is that of ambient isotopy, which is a continuous deformation of one knot into another which also deforms the ambient space continuously. For smooth knots this is equivalent to one being smoothly deformable into another by a deformation applied just to the knots themselves, not acting on the ambient space. Every polygonal knot is equivalent to a smooth knot and every smooth knot is equivalent to a polygonal knot, and a knot is called tame if it is equivalent to a polygonal knot, or (equivalently) if it is equivalent to a smooth knot.

Alexander works quite explicitly with polygonal knots, whereas Jones phrases his version of the argument for smooth knots. Ultimately these give the same conclusion, since every polygonal knot is equivalent to a smooth knot and vice versa; but the arguments are (necessarily) quite different. De Toffoli and Giardino oscillate between regarding the knot they are discussing as polygonal and as smooth as they move through the argument, following Alexander in places and Jones in others, and this is where some of the confusion stems from.⁴ When discussing De Toffoli and Giardino's paper, Larvor notes that Alexander did work with polygonal knots, and wonders whether this might matter to their conclusions (Larvor 2019, p. 2728); he is right to wonder about this, though does not fully realise its importance.

Now onto the result itself. De Toffoli and Giardino phrase this as showing that any knot is equivalent to a closed braid (see their paper for an account of braids, whose nature will not be important here). They limit themselves to arguing for the result seen in Sect. 2, that any tame knot has a diagram in which there is an axis around which the knot always goes the same way—always clockwise or always anti-clockwise. This is Alexander's original lemma, which had no mention of braids—braids were only defined a few years later—though the fact that every tame knot has a representation as a closed braid is a quick corollary.

The relevant part of De Toffoli and Giardino's account of the argument (De Toffoli and Giardino 2016, p. 41) starts by taking a tame knot K with diagram \mathcal{D}_K , and taking this diagram \mathcal{D}_K to be polygonal—thus implicitly assuming that K is polygonal (which they can do since any tame knot is equivalent to a polygonal knot). They take a small linear piece AB of \mathcal{D}_K which does not contain more than one crossing, and choose a point C such that O lies in the triangle ABC . They replace AB in the diagram by the two segments AC and CB . This gives a precise description of the intended modification to the knot diagram, but leaves open the question of what the modification of the knot K is which leads to this change in \mathcal{D}_K . This is one of the key points where De Toffoli and Giardino depart from Alexander's proof.

To explain what modification of K gives rise to this change in \mathcal{D}_K , they appeal to a Jones's version of the argument. He is working with smooth knots, rather

⁴ It is clear that De Toffoli and Giardino are aware though that smooth and polygonal knots are different, as seen for instance in their footnote 26 (De Toffoli and Giardino 2016, p. 41).

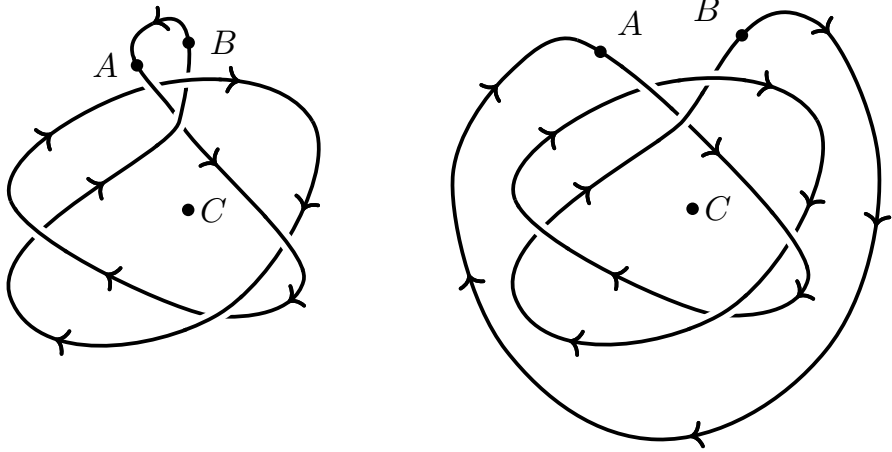


Fig. 3 The over the shoulder manoeuvre

than polygonal knots, and phrases this key part by saying that one “throws it over one’s shoulder”, referring to the short stretch of knot being focused on (Jones 1998, p. 211). He illustrates this with a diagram like that of Fig. 3. De Toffoli and Giardino repeat Jones’s phrase, saying that one throws the segment AB over one’s shoulder (Jones 1998, p. 211). They reference pictures and videos of how this manoeuvre could be carried out on a smooth knot, looking again like Fig. 3.

Basing this part of their argument on Jones’s version, their description and pictures of this manoeuvre involve smooth knots. This clashes badly with the context, as by assumption their knot is polygonal. They describe how

Intuitively, the move consists in replacing a portion of the knot that goes in the opposite direction by throwing it in the other side of the point O so that it goes in the right direction. This has to be done carefully, without introducing new entanglements. (De Toffoli and Giardino 2016, pp. 41–42)

They do not describe why care is needed, how entanglements could be introduced, or how they could be avoided—and it is appears that again in this remark they are describing the smooth rather than polygonal case.

It appears they feel that in this description they are clarifying details left implicit by Alexander, quoting Alexander as saying

the transformation of \mathcal{D}_K *obviously* corresponds to an isotopic transformation of the space curve L (De Toffoli and Giardino 2016, pp. 42, emphasis De Toffoli and Giardino, notation changed by them)

Here they use L in place of K , as Alexander is discussing a system of linked knots (for which De Toffoli and Giardino are introducing this symbol L) rather than just a single knot.

Their final remark is that by repeating the process, one can eliminate every segment of the diagram which went in the wrong direction, and obtain the desired result.

There are two key respects here in which De Toffoli and Giardino unwittingly alter Alexander's argument. One, highlighted above, is in what modification is made to the knot K that leads to the described modification of the diagram \mathcal{D}_K . The second is in the structure of the argument, leading De Toffoli and Giardino to believe that intuition is required to deliver that the process of knot modifications terminates. These alterations are the source of De Toffoli and Giardino's bolder claims about the argument, which they use as grounds for their analysis of proof more generally. The first is discussed in Sect. 4, and the second in Sect. 5.

4 The “legitimate operations”

Firstly we have the nature of the knot modification Alexander uses: given (in De Toffoli and Giardino's notation) a knot K , we make some modification to it that corresponds to the transformation of the diagram \mathcal{D}_K discussed in Sect. 3. When interpreting this De Toffoli and Giardino drawing on Jones's version of the argument, despite Jones working with smooth rather than polygonal knots, meaning De Toffoli and Giardino's description and pictures make little sense with regard to the polygonal knot K they (and more importantly, Alexander) are working with.

De Toffoli and Giardino then misinterpret Alexander's phrase of modifying the knot using “legitimate operations”. Working from their pictures and description—based entirely on Jones's version of the argument, and a recent video by Dalvit (2012)—they infer that Alexander is appealing to a shared practice amongst topologists of envisioning continuous transformations. They believe that this form of reasoning is not propositional and cannot be reduced to formal statements. They thus believe that Alexander's argument is not valid according to any general standard of validity that applies throughout mathematics, only being valid according to a special, local standard of validity (based on envisioning these kinds of continuous transformations) used in some areas of low dimensional topology. This is the main basis for their claim about mathematics being broken up into separate communities, each with their own standard of validity, that was discussed in Sect. 2. A second contributing factor to this claim is their altering the structure of Alexander's argument, discussed in Sect. 5.

To understand what Alexander actually means, it will help to make clearer the context of the relevant part of his argument. Firstly, Alexander is quite explicit that he is working with a polygonal notion of knot, assuming that a knot is composed of a finite number of straight line segments in \mathbb{R}^3 (Alexander 1923, p. 93). This is central to the way his proof works, as he moves through the finite number of straight line segments one by one, fixing any which in the diagram go the wrong way around the axis (Sect. 5 discusses the structure of his argument more closely).

When discussing Alexander's proof, I will base my notation on De Toffoli and Giardino's from Sect. 3, rather than Alexander's, but it is useful to supplement it. I will write $[a_1, \dots, a_n]$ for the convex hull of a_1, \dots, a_n , defined to be

$$[a_1, \dots, a_n] = \left\{ \sum_{i=1}^n \lambda_i a_i \mid 0 \leq \lambda_i \leq 1, \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Thus for instance $[a, b]$ is the line segment between point a and point b (for a, b distinct), and $[a, b, c]$ is the closed triangular region with a, b, c as its vertices (for a, b, c not collinear). If $a < b \in \mathbb{R}$ then this segment $[a, b]$ is the usual closed interval with endpoints a and b .

We have a polygonal knot K in \mathbb{R}^3 , with projection \mathcal{D}_K onto a plane P . Let $\pi : \mathbb{R}^3 \rightarrow P$ be the orthogonal projection, so $\mathcal{D}_K = \pi(K)$. We have a picked a point O in P , and we are modifying \mathcal{D}_K so that it only goes clockwise (say) around O . $[A, B]$ is a subsegment of \mathcal{D}_K which goes anti-clockwise, and such that \mathcal{D}_K has at most one crossing on $[A, B]$. We select a point C such that the point O lies in the interior of the triangle $[A, B, C]$. We seek to find a knot K' which is equivalent to K such that the diagram $\mathcal{D}_{K'}$ of K' is the same as \mathcal{D}_K , but with the two segments $[A, C]$, $[C, B]$ replacing the single segment $[A, B]$. This is the context of the quote from Alexander seen at the end of Sect. 3:

The transformation of \mathcal{D}_K obviously corresponds to an isotopic transformation of the space figure K . (Alexander 1923, p. 94)

(the notation here has been modified to fit with De Toffoli and Giardino's account⁵).

This is where De Toffoli and Giardino appeal to Jones's version of the argument, for the smooth case, using his phrase about throwing the knot over one's shoulder, with a diagram like that in Fig. 3. They also use stills from a video by Dalvit (2012) made to illustrate the smooth version of the argument. As mentioned in Sect. 3 and at the start of this section, this makes little sense in the context Alexander is working. His knot is polygonal and no smooth isotopy can be applied to it (due to kinks in the knot where the different segments meet). Also, the kinds of continuous/smooth transformations that De Toffoli and Giardino describe and picture would not lead to a result with the required diagram—the same as that of K , but with the two segments $[A, C]$, $[C, B]$ replacing the single segment $[A, B]$. If one isotoped K into a smooth knot, the result would have a smooth diagram, not a polygonal diagram.

However if one puts aside Jones's version of the argument and instead focuses just on what Alexander is saying, it is clear what he means. We will suppose first that there is a single segment of K lying above $[A, B]$, so that there are $a, b \in K$ such that $[a, b] \subseteq K$ and $\pi([a, b]) = [A, B]$ (actually it appears to be an oversight by Alexander that this is not guaranteed at this point, as will be discussed later in this section; a slight rephrasing of the argument would guarantee this). If \mathcal{D}_K has a crossing point on $[a, b]$, with $x \in [a, b]$ such that there is $y \notin [a, b]$ with $\pi(x) = \pi(y)$, then we can assume WLOG (by a rotation of space) that $x - y$ points vertically upwards. Thus the region vertically above the line segment $[a, b]$ is free from obstructions.

⁵ Alexander here is actually talking about a linked system of knots S , rather than a single knot K , but this is of no importance for us.

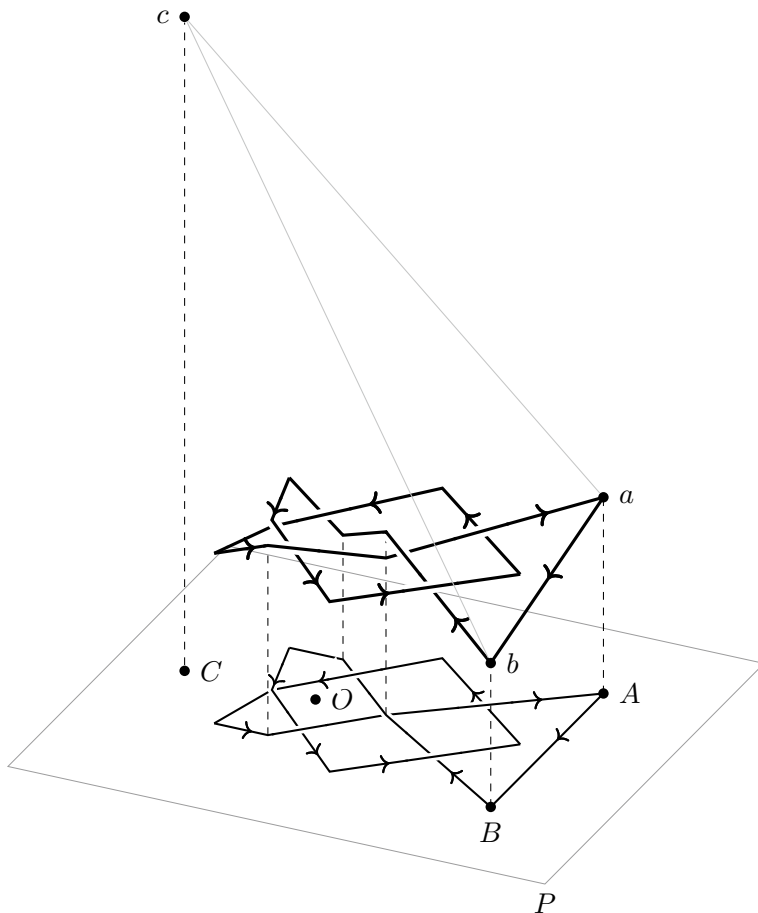


Fig. 4 Avoiding K with the triangle $[a, b, c]$

We are seeking a knot K' obtained by an isotopic transformation of K such that $\pi(K')$ is the same as $\pi(K) = \mathcal{D}_K$ but with the two segments $[A, C]$, $[C, B]$ replacing the single segment $[A, B]$. Thus K' must have the line segment $[a, b]$ replaced by some combination of line segments in \mathbb{R}^3 whose projection (under π) is $[A, C] \cup [C, B]$. So there must be a point c with $\pi(c) = C$, and a joined to c in K' by a sequence of line segments which project to $[A, C]$, and c joined to b in K' by a sequence of line segments which project to $[C, B]$. Does such a point c exist?

Obviously yes. As we are visualizing it, the region vertically above $[a, b]$ is free from obstructions, so if we take c to be enormously high up then the triangle $[a, b, c]$ will go almost straight up from the line segment $[a, b]$, and will not hit anywhere in K —in other words, with $[a, b, c] \cap K = [a, b]$. This is illustrated in Fig. 4. Thus we can take K' to be K but with $[a, b]$ replaced by $[a, c] \cup [c, b]$, which has the required projection, as seen in Fig. 5.

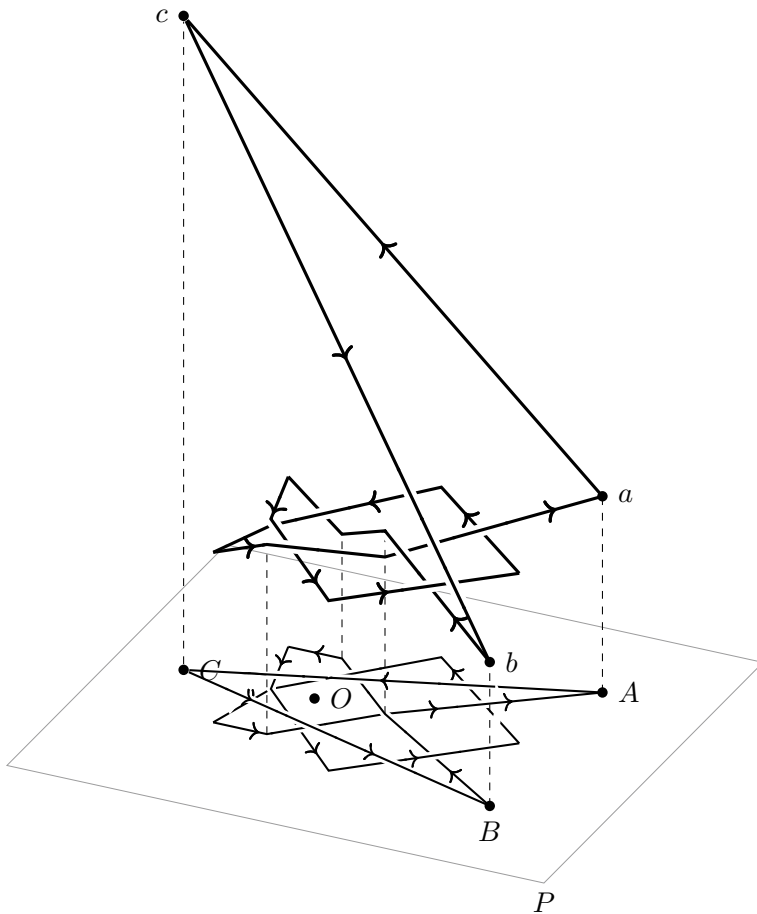


Fig. 5 The result of replacing $[a, b]$ with $[a, c] \cup [c, b]$

There is no question that this is what Alexander intends, rather than the vaguely specified continuous/smooth transformation De Toffoli and Giardino describe and picture. Perhaps they were attempting to make the proof more accessible to a lay audience, but in truth their account is more complex and confusing than the simple pictures in Figs. 4 and 5.

It is clear from the preceding Alexander has a notion of isotopy in mind on which if we have a knot K with a segment $[a, b]$ and a point c such that the triangle $[a, b, c] \cap K = [a, b]$, then K is isotopic to K' where K' is the same as K but with $[a, b]$ replaced by $[a, c] \cup [c, b]$. If Alexander had a notion of isotopy in mind on which this was not possible, his paper would be misleading at this key point. We don't need to know any more about his notion of isotopy than this to follow his argument, and this much we can infer from it.

It turns out that this is essentially exactly the standard notion of equivalence for polygonal knots. Alexander gives the definition in another paper:

On any edge AB we may construct a triangle ABC , so drawn that neither the vertex C , the edge AC , the edge CB , nor the plane triangular region bounded by ABC has a point in common with the knot. We may then transform the knot by removing the edge AB and substituting in its place the edges AC and CB , along with the vertex C . We may also perform the reverse operation which consists in replacing a pair of consecutive edges AC and CB , together with their common vertex C by a single edge AB , provided neither the edge AB nor the plane triangular region bounded by ABC has a point in common with the knot. Each of the transformations here described will be called an elementary deformation. (Alexander and Briggs 1926, p. 563)

He defines two knots K_1 and K_2 to be of the same type if they can be related by a finite sequence of elementary deformations of the above kind. If this holds I will instead say that K_1 can be polygonally deformed into K_2 . This is an equivalence relation.

The standard notion of equivalence for arbitrary knots (not just polygonal) is that of ambient isotopy. We define a knot here to be a continuous injective map $\phi : S^1 \rightarrow \mathbb{R}^3$. Then an ambient isotopy is a continuous map $H : \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$ such that $t \mapsto H(t, s)$ is a homeomorphism $\mathbb{R}^3 \rightarrow \mathbb{R}^3$ for all s and $H(t, 0) = t$ for all t . If ϕ, ψ are knots, an ambient isotopy from ϕ to ψ is an ambient isotopy H such that $H(\phi(t), 1) = \psi(t)$ for all t . We call ϕ, ψ ambient isotopic if an ambient isotopy from ϕ to ψ exists, and this is an equivalent relation on knots.

It is in fact the case that two polygonal knots are equivalent under polygonal deformation iff they are ambient isotopic. This is a basic fact of knot theory, in a sense more basic than the equivalence of smooth and piecewise linear notions of knot that De Toffoli and Giardino cite (De Toffoli and Giardino 2016, p. 41, footnote 26). One direction of this equivalence of equivalences is easy: that if K_1 and K_2 are polygonal knots such that K_1 can be polygonally deformed into K_2 , then K_1 is ambient isotopic to K_2 (actually ambient isotopic via a piecewise linear ambient isotopy). This is proved for instance as one of the first propositions in Burde and Zieschang (2002, pp. 6–7, implication (3) \Rightarrow (2) of Proposition 1.10). Thus under either polygonal deformation or ambient isotopy, it is clear that replacing $[a, b]$ in K by $[a, c] \cup [c, b]$ gives an equivalent knot (the former by definition, the latter by a simple argument). Thus under either definition Alexander’s proof is valid, and we do not need to know which one he intended to follow it.

I will shortly discuss how De Toffoli and Giardino’s claims hold up in light of all these points. Before that there are two things that should be remarked on. The first is the existence of a point c high enough up that $K \cap [a, b, c] = [a, b]$. This is a good example of the kind of high level reasoning discussed in Sect. 1. Someone trained in maths can “see” this to be true by visualizing the situation; but it is also clear how one would spell this out in greater detail. For $\lambda \geq 0$ let $c_\lambda = c + \lambda n$ where n is the normal to P pointing “upwards”, i.e. in the direction of $x - y$ if K has a crossing point $x \in [a, b]$ with $\pi(y) = \pi(x)$, $y \notin [a, b]$, as discussed above (if there is no such x we can take n to be any non zero normal to P). Then the claim is that for λ

sufficiently large, $K \cap [a, b, c_\lambda] = [a, b]$. We can split this up into multiple subclaims. Let $[d, a]$ be the edge of K preceding $[a, b]$, and $[b, e]$ the edge following $[a, b]$. Let $[p, q]$ be the edge containing y if there is such a y , otherwise we can take $[p, q] = \emptyset$. Then we have that

$$K \setminus ((d, a] \cup [a, b] \cup [b, e) \cup (p, q))$$

is compact with

$$\pi(K \setminus ((d, a] \cup [a, b] \cup [b, e) \cup (p, q))) \cap [A, B] = \emptyset,$$

and we need to argue that:

- For λ sufficiently large, $[a, b, c_\lambda] \cap [d, a] = \{a\}$
- For λ sufficiently large, $[a, b, c_\lambda] \cap [b, e] = \{b\}$
- For λ sufficiently large, $[a, b, c_\lambda] \cap [p, q] = \emptyset$
- For λ sufficiently large, $[a, b, c_\lambda] \cap (K \setminus ((d, a] \cup [a, b] \cup [b, e) \cup (p, q))) = \emptyset$.

Each of these can indeed be proved in greater detail if necessary. It appears that this comes to a few pages, if written out comprehensively. Of course one does not have to write this out to see Alexander's proof to be valid; but it is important for rigour that it be possible to argue the inference in greater detail if called for, and that it is not an irreducibly high level intuition. As discussed in Sects. 1 and 2, I think there is an important question for the epistemology of mathematical proof here: how and in what circumstances can one gain the ability to reliably judge high level inferences like this to be provable in greater detail?

The second point is that in the above, I introduced the assumption that there is a single segment of K lying above $[A, B]$ —that there are $a, b \in K$ such that $[a, b] \subseteq K$ and $\pi([a, b]) = [A, B]$. If one was careless when visualizing the situation one might well assume that $[A, B]$ would have to have such a line segment $[a, b]$ lying above it, but in fact this need not be the case: all we can guarantee is that there is a finite sequence $[a_1, b_1], \dots, [a_n, b_n]$ of line segments contained in K with $[A, B] = \bigcup_{i=1}^n \pi([a_i, b_i])$. Each of these line segments $[a_i, b_i]$ must lie above the line segment $[A, B]$, but they can have different vertical components to their gradients. In this case the argument proceeds much the same way as above, but the point c has to be picked high enough that for each i , the triangle $[a_i, b_i, c]$ only intersects K in $[a_i, b_i]$. The unnecessary complication this creates appears to be a simple oversight by Alexander. When he talks about “ P mov[ing] along certain segments of the broken line” (Alexander 1923, p. 94) he could just as easily talk instead about P moving along the projection of certain segments of the knot above. This would not affect the rest of his proof at all, and in this case each segment $[A, B]$ like the one we considered above would have a single segment $[a, b]$ of K above it.

Now to De Toffoli and Giardino's claims about this part of the argument. First, they claim that the reasoning is not propositional reasoning, nor formal reasoning, and is not based on formal reasoning, nor can it be reduced to formal statements (De Toffoli and Giardino 2016, pp. 43–44, 48–49). It is not entirely clear what they mean by this. Alexander's written proof consists entirely of words and symbols, and

contains no pictures—in what sense is it “not propositional”? One can reason from a first proposition to a second in many different ways, including via one’s spatiotemporal faculties. Perhaps when they say propositional reasoning, they mean reasoning in terms of strict logical rules; and of course Alexander’s argument is not literally a formal argument—nor are most published proofs. This is not a significant point though, and I doubt anyone has ever claimed the opposite. Although Alexander’s proof is not formal, as discussed in Sect. 1 and above it is important for rigour that its inferences be provable in greater detail if requested; and this is indeed the case, as sketched for one key inference above. If one keeps repeating this process, asking for greater detail/more precision in every inference, and then for greater detail/more precision in each of those more detailed inferences in turn, one will eventually reach a formal proof. This is line with the briefly sketched argument in Sect. 1 that all rigorous proofs are formalizable, as a consequence of the norm of rigour. I am not claiming any epistemic benefits to this here however, just noting that it can be done.

With regard to the claimed importance of non-propositional reasoning, it is also worth noting that the crucial clarifications of Alexander’s argument given above were propositional—the correct intended knot modification, the existence of the point c high enough above the knot that $K \cap [a, b, c] = [a, b]$, and so on. These propositions can be illustrated visually, but if one had to limit oneself to the propositions or the illustrations in writing out the argument, I think the propositions would be the part to keep.

Although Alexander’s proof would require a normal mathematician to do some visualising to follow it, De Toffoli and Giardino do not quite grasp the nature of the visualization involved. They describe Alexander’s proof as based on the manipulation of concrete spatio-temporal objects (De Toffoli and Giardino 2016, p. 44), which is inaccurate as Alexander’s proof is based on knots being a finite union of straight line segments, which are not concrete and have zero width (of course in some crude sense one would could trace back a grasp of how straight lines behave to familiarity with concrete objects, but in this sense almost all mathematical reasoning would be based on the concrete and the claim is uninteresting). They repeatedly refer to Alexander’s argument as involving smooth or continuous transformations (De Toffoli and Giardino 2016, pp. 41, 43, 44, 45, 46). As discussed above, Alexander intends a polygonal deformation of the knot; referring to this as “continuous” is misleading in its excess generality, and referring to it as “smooth” is incorrect. This polygonal deformation requires a much more straightforward visualization than the continuous/smooth ones they indicate in their various diagrams (De Toffoli and Giardino 2016, pp. 42). Their remarks about being careful not to introduce new entanglements while transforming the knot might be pertinent to Jones’s version, but are not relevant to Alexander’s actual proof with its simple polygonal transformation (De Toffoli and Giardino 2016, p. 42).

This all leads them to overestimate the role played by visualization in the proof, which is much simpler and more easily backed up by detailed arguments than they describe. This completely undermines their claim that Alexander’s proof relies on a special “local” standard of validity used by topologists, in terms of envisioning continuous transformations (De Toffoli and Giardino 2016, pp. 43–46, 48–49). In fact Alexander’s proof is perfectly rigorous by the usual standards in mathematics

(with a mild imperfection in the point noted above that he should guarantee a single segment of K lying above $[A, B]$, but does not, unnecessarily complicating things slightly).

De Toffoli and Giardino's claims here are really only suited to Jones's version of the argument—they are right that Jones's version seems to have fairly irreducible appeals to intuition and spatiotemporal reasoning, and that it would be very difficult to prove in greater detail or formalize. This is not of so much interest however as Jones's argument is not a published mathematical proof, but a short sketch of a proof in a paper consisting of philosophical musings. I discuss Jones's version of the argument, for the smooth case, in another paper,⁶ arguing that it is not remotely rigorous by normal standards—and thus that it also presents no challenge to the view of rigour sketched in Sect. 1.

5 Termination of the Process

There a second respect in which De Toffoli and Giardino misrepresent Alexander's argument which leads them to overstate its reliance on visualization and intuition. The proof describes a sequence of modifications to a knot, and it is essential to the proof that this sequence eventually terminates, in a knot with a diagram of the required form (only going the right way around an axis in the plane); if it does not terminate, the lemma fails. Here De Toffoli and Giardino claim that

it is left to our intuition to prove that ...it is not an infinite process. Alexander does not really [give] us any other justification: this reasoning plays an epistemic role. (De Toffoli and Giardino 2016, p. 44)

However as was the case in Sect. 4, their conclusion rests on a confusion. In this case, they miss out key steps in Alexander's reasoning, which ensure the termination of the process. They are wrong to think that in their version of the argument "intuition" could guarantee the termination of the process—in the argument as they have stated it, there is no guarantee that the process will terminate.

I will start by discussing the problem with De Toffoli and Giardino's version of the argument. They choose a small straight portion $[A, B]$ of the diagram, which goes the wrong way around O and contains at most one crossing, and they correct this one segment—bending it to go the other way around O . They then move onto another small straight portion of the diagram which goes the wrong way and contains at most one crossing, and do the same. Since the diagram has only finitely many crossings, one might hope that this process would always terminate. The problem is that when bending a segment to go the right way, one may introduce *extra* crossings to the diagram, and one may in fact introduce extra crossings to the troublesome parts of the diagram—that go the wrong way around O . Thus one could potentially keep on going forever, bending more and more segments of the diagram to go the right way, but constantly adding to the workload as one goes

⁶ "Rigour, Pictures and Knot Theory", under review.

by increasing the number of troublesome crossings. The diagram would get more and more complicated, with smaller and smaller segments being bent the right way each time. The lemma would fail.

De Toffoli and Giardino (2016, p. 44) show some awareness of this problem in the above quote, but are wrong to think that it can be brushed aside by “intuition”. In the process as they describe it, they have left the above possibility wide open. It would not be difficult to describe a sequence of knot modifications that fits their description but never terminates: take two troublesome sections S_1 and S_2 on opposite sides of O , and first correct a section of S_1 containing a crossing while simultaneously adding at least one troublesome crossing to S_2 , then correct a section of S_2 containing a crossing while simultaneously adding at least one troublesome crossing to S_1 , and so on. Of course one could use one’s “intuition” to see that this could be avoided—that one could give a more careful description of the process that ruled out this possibility. But that is not to use intuition to see their argument is valid: it would be to use intuition to rewrite their argument to make it valid. Their comment that one has to carry out the over the shoulder manoeuvre “carefully” to avoid introducing new entanglements (De Toffoli and Giardino 2016, p. 42) does not help, since the procedure being described is one that has to work without human oversight or intelligence (it has to work just as well for a knot with 10^{1000} crossings as with 10).

In fact, the problem is easily avoided, as seen in Alexander’s actual proof. The key difference between his proof and the version described by De Toffoli and Giardino is in its logical structure—exactly the kind of feature that a perspective focused overmuch on visualization and intuition is likely to miss. Alexander’s proof is not an induction, which is the attempted structure of De Toffoli and Giardino’s; it is a *double* induction, with the part of the argument described by De Toffoli and Giardino being the inner induction.

In fact, Alexander’s proof first considers the set of segments of \mathcal{D}_K which bend the wrong way around O (in his notation, he considers the set of segments of S_π which bend the wrong way around L). I will call this set T here. His argument deals with each element σ of T in turn, by breaking each such σ up into finitely many subsegments $\sigma_1, \dots, \sigma_n$ on each of which there is at most one crossing. The point is that one when one corrects the subsegment σ_i one does not add crossings to σ —though one may add crossings to other elements of T . To make this completely clear, we can phrase the argument as follows. I will not be entirely formal here, sufficing to make clear this double induction structure.

Proposition *Suppose K is a polygonal knot and σ a line segment contained in \mathcal{D}_K which goes around O the wrong way. Suppose σ_i is a subsegment of σ such that \mathcal{D}_K has at most one crossing on σ . Then K is equivalent to a polygonal knot L with the same diagram as K , except with the subsegment $\sigma_i = [A, B]$ replaced by two segments $[A, C]$ and $[C, B]$ with C a point such that $O \in [A, B, C]$.*

Proof This is the part of the argument discussed in Sect. 4, and the part that appears in De Toffoli and Giardino's account (in somewhat altered form, as discussed in Sects. 3 and 4). \square

Proposition *Suppose K is a polygonal knot and σ a line segment contained in \mathcal{D}_K which goes around O the wrong way. Then K is equivalent to a polygonal knot L which has the same diagram as K outside of σ , and such that L 's diagram goes the right way around O on the part it replaces σ with.*

Proof We break σ up into subsegments $\sigma_1, \dots, \sigma_n$ such that each σ_i has at most one crossing. Then by repeatedly applying the previous proposition (this is the inner induction) to each σ_i in turn, we obtain the result. Here we use the fact that if $\sigma_i = [A, B]$ and C is a point such that $O \in [A, B, C]$, then $([A, C] \cup [C, B]) \cap \sigma = \{A, B\}$, so that replacing σ_i with $[A, C] \cup [C, B]$ does not add any crossings to any σ_j for $j > i$. \square

Proposition *Suppose K is a polygonal knot. Then K is equivalent to a polygonal knot L with a diagram which only goes around O the right way.*

Proof This is by induction on the size of the set of segments of \mathcal{D}_K which go around O the wrong way, with the previous proposition providing the induction hypothesis (and the base case trivial). \square

Thus Alexander's argument here is perfectly rigorous—by the normal standards—as stated. The apparent flaw De Toffoli and Giardino discuss, the possibility that the process need not terminate—which they look to intuition to solve—is a flaw their version inherits from Jones's, and has no root in Alexander's original argument.

6 Conclusion

The brief account of rigour sketched in Sect. 1 is unthreatened by Alexander's proof. On the contrary, Alexander's proof is a good illustration of it. All of De Toffoli and Giardino's stronger claims about Alexander's argument rest on two alterations: concerning the nature of the knot deformation Alexander intends, and the structure of his argument. With these points cleared up, their claims about his argument are seen to have no basis. This in turn removes the grounds for their more general claims about mathematics being split into different communities, each with their own standard of validity, claims which were critiqued in Sect. 2.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abbott, S. (2016). *Understanding analysis*. Berlin: Springer.
- Alexander, J. W. (1923). A lemma on systems of knotted curves. *Proceedings of the National Academy of Sciences of the United States of America*, 9(3), 93–95.
- Alexander, J. W., & Briggs, G. B. (1926). On types of knotted curves. *Annals of Mathematics*, 28(1), 562–586.
- Antonutti Marfori, M. (2010). Informal proofs and mathematical rigour. *Studia Logica: An International Journal for Symbolic Logic*, 96(2), 261–272.
- Azzouni, J. (2013). The relationship of derivations in artificial languages to ordinary rigorous mathematical proof. *Philosophia Mathematica*, 21(2), 247–254.
- Burde, G., & Zieschang, H. (2002). *Knots*. New York: de Gruyter.
- Burgess, J. P. (2015). *Rigor and structure* (p. 240). Oxford: Oxford University Press.
- Celluci, C. (2009). Why proof? What is a proof? In R. Lupacchini & G. Corsi (Eds.), *Deduction, computation, experiment* (pp. 1–27). Berlin: Springer.
- Dalvit, E. (2012). A journey through the mathematical theory of braids. <http://matematita.science.unitn.it/braids/>
- De Toffoli, S., & Giardino, V. (2016). Envisioning transformations—The practice of topology. In B. Larvor (Ed.), *Mathematical cultures: The London Meetings 2012–2014* (pp. 25–50). New York, NY: Birkhäuser.
- Dunham, W. (1999). *Euler: The master of us all* (p. 208). Washington, DC: The Mathematical Association of America.
- Goethe, N. B., & Friend, M. (2010). Confronting ideals of proof with the ways of proving of the research mathematician. *Studia Logica*, 96(2), 273–288.
- Gowers, T., Barrow-Green, J., & Leader, I. (Eds.). (2008). *The Princeton companion to mathematics* (p. 1056). Princeton: Princeton University Press.
- Hirsch, M. W. (1976). *Differential topology*. Graduate texts in mathematics New York: Springer.
- Jones, V. (1998). A credo of sorts. In H. G. Dales & G. Oliveri (Eds.), *Truth in mathematics* (pp. 203–214). Oxford: Clarendon Press.
- Larvor, B. (2012). How to think about informal proofs. *Synthese*, 187(2), 715–730.
- Larvor, B. (2019). From Euclidean geometry to knots and nets. *Synthese*, 196(7), 2715–2736.
- Lee, J. (2012). *Introduction to smooth manifolds*. Graduate texts in mathematics New York: Springer.
- Leitgeb, H. (2009). On formal and informal provability. In O. Bueno & Ø. Linnebo (Eds.), *New waves in philosophy of mathematics* (pp. 263–299). Basingstoke: Palgrave Macmillan.
- Pelc, A. (2009). Why do we believe theorems? *Philosophia Mathematica*, 17(1), 84–94.
- Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, 7(1), 5–41.
- Rav, Y. (2007). A critique of a formalist–mechanist version of the justification of arguments in mathematicians’ proof practices. *Philosophia Mathematica*, 15(3), 291–320.
- Reidemeister, K. (1927). Elementare Begründung der Knotentheorie. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 5(1), 24–32.
- Weir, A. (2016). Informal proof, formal proof, formalism. *The Review of Symbolic Logic*, 9(1), 23–43.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.